

V3 Loop Amino Acids

The following pages present amino acid alignments of the V3 loop, arranged by phylogenetic subtype. For each subtype, the number of sequences used to construct the alignment is indicated. The top line in each alignment represents the consensus sequence for that subtype, where “consensus” simply means the most common amino acid found in each position among the sequences of the given subtype. The subscripts record the frequency with which that amino acid is observed at that location among members of the subtype. An amino acid which is conserved 100% is shown with no subscript. Directly beneath the most common amino acid in each position are the other amino acids observed in that position, listed from most common to least common. An asterisk (*) subscript means less than 0.5% of the sequences had the indicated amino acid at that location. A dash (-) indicates a gap inserted to maintain the alignment. Percentages were rounded to the nearest whole number.

For this year's alignment, the HMMER (version 1.8) hidden Markov model software (Sean Eddy, Dept. of Genetics, Washington U. School of Medicine, St. Louis, MO 63110; eddy@genetics.wustl.edu <http://hmmer.wustl.edu/>) was used to objectively align all 3,826 sequences. The frequency counts are derived from this alignment.

Both the HMMER alignment of nucleotide sequences, and an edited version translated to amino acid sequences, will be available from the LANL HIV database (<http://hiv-web.lanl.gov>). Questions about these alignments should be directed to Brian Foley btf@t10.lanl.gov (505) 665-1970.

A subtype (539 sequences)

B subtype (1912 sequences)

C subtype (443 sequences)

D subtype (182 sequences)

CRF01(AE) form (356 sequences)

N₉₆ C₉₉ T₉₇ R₉₉ P₁₀₀ S₈₇ N₈₉ N₉₃ T₈₈ R₉₅ T₉₂ S₉₀ I₇₅ T₆₃ I₈₅ G₁₀₀ Q₉₆ V₈₈ F₉₃ Y₉₇ R₈₀ T₉₇ G₉₆ D₈₀ I₉₇ L₇₂ G₁₀₀ D₈₅ I₉₅ R₉₈ K₈₅ A₉₉ Y₉₂ C₉₉ E₉₂
 S₃ Y₁ I₂ I₁
 K₁ F₁₀ K₄ K₃ R₄ M₈ R₁₅ M₁₀ E_{*} R₁₀ A₄ L₃ H₂ K₁₆ P₁ E₃ V₁ T₁₈ R_{*} N₁₄ P₂ K₁ Q₆ S₁ F₅ S₁ K₃
 S₁ K_{*}
 T₁ P_{*}
 D_{*}
 H_{*}

N₁ T₄ R₁ K₄ I₆ K₄ I₂ K₃ E_{*} R₁ K₁ T₄ M₂ E_{*} R₁ K₁ T₄ P₅ L₁
 T₁ I₁ Y₁ T₁ R₁ K₁ P₁ L₃ H₄ R₁
 Y₁ D₁ T₁ R₁ S_{*} A₁ P₁ L₁ A₂
 A_{*} G₁ I₁ V₁ * P_{*} F₁ A₃
 R₁ S_{*} P_{*} Q_{*} E_{*} S₁ K_{*}
 E_{*} Y_{*} G_{*} V_{*}
 A_{*} Y_{*} S_{*} P_{*} E_{*} E_{*} G_{*}
 S_{*}

F subtype (84 sequences)

N₈₉ C₁₀₀ T₉₄ R₉₉ P₁₀₀ N₉₆ N₉₈ T₉₆ R₁₀₀ K₉₃ S₈₈ I₉₉ H₆₂ L₆₀ G₉₉ P₉₉ G₉₉ Q₆₀ A₈₈ F₉₉ Y₉₈ A₆₂ T₉₈ G₉₉ D₅₅ I₉₉ I₉₈ G₁₀₀ D₉₅ I₉₅ R₁₀₀ K₈₇ A₉₉ H₈₈ C₉₉ N₈₈
 I₆ S₂ S₁
 D₁ R₁
 T₂ I₁
 P₁

S₄ K₁ T₁ I₂
 S₁ Y₁ K₁
 E₁ R₆
 R₁₁ F₆
 P₄ V₁
 S₂
 K₁
 Y₁

R₆ G₁₂ I₃₃ A₁ L₁ R₁ R₄₀ T₆ L₁ H₁ T₃₈ A₂ D₁ A₂
 V₅ F₁
 K₁
 S₁
 G₁
 I₁

L₆₀ G₉₉ Q₆₀ A₈₈ F₉₉ Y₉₈ A₆₂ T₉₈ G₉₉ D₅₅ I₉₉ I₉₈ G₁₀₀ D₉₅ I₉₅ R₁₀₀ K₈₇ A₉₉ H₈₈ C₉₉ N₈₈
 E₂ T₂
 R_s S₁
 N₁ L₁
 E₁ P₂
 D₁ S₁
 K₁ D₁

G subtype (73 sequences)

N₄₂ C₁₀₀ T₉₀ R₉₉ P₉₉ N₈₄ N₉₉ N₁₀₀ T₉₉ R₉₅ K₉₀ S₉₂ I₉₆ R₂₅ I₃₈ * G₉₀ P₉₇ G₉₉ Q₉₇ A₇₉ F₇₅ Y₉₉ A₈₂ T₉₇ G₉₂ D₃₈ I₉₇ I₉₇ G₁₀₀ D₇₈ I₉₉ R₁₀₀ Q₉₀ A₁₀₀ H₈₈ C₁₀₀ N₉₅
 I₂₁ S₃ V₃
 R₅
 V₅
 M₄
 S₃
 H₁

K₁ F₁ S₁₀ D₁
 G₄
 H₁

K₁ M₄ R₅ V₃
 I₁ G₁ K₁ H₁₈ L₂₅
 I₁ T₁₈ M₄
 P₁₁
 N₄
 S₃
 H₁
 E₁
 K₁

P₁
 R₃
 I₁
 V₄
 S₁
 T₁
 D₃
 A₁₂
 T₁
 N₈
 G₃
 L₁

H subtype (11 sequences)

N₅₅ C₁₀₀ T₃₂ R₁₀₀ P₈₂ N₆₄ N₁₀₀ N₁₀₀ T₉₁ R₉₁ K₆₄ S₈₂ I₉₁ H₆₄ L₇₃
T₄₅ I₉ L₉ G₂₇ A₉ T₉ A₉ P₉₁ G₁₀₀ Q₇₃ A₉₁ F₇₃ Y₈₂ A₃₂ T₇₃ G₆₄ D₅₂ I₈₂
I₁₀₀ G₁₀₀ D₇₃ I₁₀₀ R₁₀₀ Q₇₃ A₁₀₀ H₈₂ C₁₀₀ N₁₀₀
L₉ E₉ R₁₈ G₁₈ M₉ R₉ G₉ I₉ A₉ R₉ R₂₇ T₉ I₁₈ H₉ T₁₈ E₉ E₁₈ V₁₈
P₉ L₉ E₉ F₉ I₉ N₉ A₉ R₉ S₉ M₉ T₉

J subtype (3 sequences)

V₆₇ C₁₀₀ T₃₃ R₁₀₀ P₁₀₀ N₆₇ N₁₀₀ N₁₀₀ T₁₀₀ K₁₀₀ R₁₀₀ G₁₀₀ M₆₇ G₁₀₀ P₁₀₀ G₁₀₀ K₄₀ F₈₀ Y₁₀₀ A₈₀ T₁₀₀ G₈₀ D₆₀ I₁₀₀ I₁₀₀ G₁₀₀ E₆₇
E₃₃ V₃₃ A₃₃ Y₃₃ R₂₀ G₄₀ R₄₀ M₂₀ R₄₀ V₄₀ V₂₀ Q₂₀ T₂₀ D₂₀ A₂₀ K₂₀
S₂₀

K subtype (5 sequences)

N₆₀ C₁₀₀ T₈₀ R₁₀₀ P₁₀₀ N₆₀ N₁₀₀ N₁₀₀ T₁₀₀ R₁₀₀ K₈₀ S₆₀ I₁₀₀ H₆₀ I₈₀ G₁₀₀ P₁₀₀ G₁₀₀ K₄₀ F₈₀ Y₁₀₀ A₈₀ T₁₀₀ G₈₀ D₆₀ I₁₀₀ I₁₀₀ G₁₀₀ E₆₇
T₂₀ I₂₀ G₂₀ A₂₀ Y₂₀ R₂₀ G₄₀ R₄₀ M₂₀ R₄₀ V₄₀ V₂₀ Q₂₀ T₂₀ D₂₀ A₂₀ K₂₀

CRF02(AG) form (57 sequences)

N₈₉ C₁₀₀ T₈₁ R₁₀₀ P₁₀₀ N₄₆ N₉₆ N₉₈ T₉₅ R₉₅ K₃₆ S₈₄ V₈₁ R₇₉ I₉₆ G₁₀₀ P₁₀₀ G₁₀₀ Q₉₆ T₈₂ F₁₀₀ Y₉₈ A₁₀₀ T₇₇ G₉₄ D₇₅ I₁₀₀ I₉₆ G₁₀₀ D₈₂ I₁₀₀ R₉₈ Q₉₁ A₁₀₀ H₉₆ C₁₀₀ N₉₁
T₅ I₁₁ G₃₉ K₂ Y₂ R₄ N₄ Q₇ G₁₂ I₁₈ H₁₈ M₂ K₂ A₁₆ L₂
I₂ S₄ S₁₄ D₂ K₂ S₂ R₄ L₂ P₄ T₂ R₂ V₂ R₉ D₄ R₅
S₂ V₄ Y₂ T₄ H₂ K₄ A₂ G₄
K₂ P₂ N₂

CRF03(AB) form (42 sequences)

N₁₀₀ C₁₀₀ T₉₈ R₁₀₀ P₁₀₀ N₁₀₀ N₁₀₀ T₁₀₀ R₉₈ K₁₀₀ G₁₀₀ I₁₀₀ H₁₀₀ I₉₈ G₁₀₀ P₁₀₀ G₁₀₀ I₉₅ T₉₀ G₁₀₀ D₉₈ I₁₀₀ I₉₅ G₉₈ D₁₀₀ I₉₅ T₉₀ G₁₀₀ D₉₈ I₁₀₀ R₉₈ Q₁₀₀ A₁₀₀ H₉₀ C₁₀₀ N₁₀₀
X₂ T₂ G₂ I₅ R₂ M₅ E₂ V₂ P₂ K₂ Y₇ L₂

CRF04(cpx) form (3 sequences)

$$\begin{aligned}
& N_{100} C_{100} T_{100} R_{67} P_{67} N_{100} N_{100} T_{100} R_{67} K_{67} S_{67} E_{33} G_{33} L_{33} \\
& G_{33} L_{33} G_{33} \\
& G_{33} T_{100} R_{67} P_{67} N_{100} N_{100} T_{100} R_{67} H_{67} I_{67} \tilde{G}_{100} P_{100} G_{100} H_{33} T_{100} G_{100} Y_{67} A_{100} I_{100} G_{100} D_{100} I_{100} R_{100} Q_{67} A_{100} H_{100} C_{100} N_{100} \\
& K_{33} F_{33} V_{33} \\
& K_{33} V_{33}
\end{aligned}$$

O group (19 sequences)